Telco Customer Churn Prediction

1. Executive Summary & Problem Framing

- **The Business Problem:** The telecom company faces a 26.5% churn rate. The goal is to identify *who* is leaving and *why*, to prioritize retention efforts.
- **The Metric that Matters:** We prioritized **Recall** (catching as many churners as possible) and **AUC-ROC** (probability ranking accuracy). Missing a churner (False Negative) is costlier than annoying a loyal customer with a retention offer (False Positive).

---

2. Data Preparation & Advanced Feature Engineering

We didn't just use raw columns; we created "Behavioral Features" to capture the *human* element behind the data.

- **TotalServices:** Count of add-ons (Security, Backup, etc.). *Hypothesis: More services = Higher switching costs.*
- **AvgCostPerService:** Monthly Charges / Total Services. *Hypothesis: Customers paying high fees for few services feel "ripped off."*
- **Risk_FiberM2M:** A specific flag for "Fiber Optic" users on "Month-to-Month" contracts. This group is notoriously volatile.

---

3. Exploratory Data Analysis (EDA)

- **The "Stickiness" Insight:** As shown in the "Probability of Churn by Number of Services" plot, churn drops dramatically as customers add more services. A customer with 0-1 services has a >40% chance of leaving. A customer with 4+ services has <10% chance.
- **Contract Risk:** Month-to-month contracts are the single biggest predictor of churn.
- **Tenure Cliff:** The "Tenure Density" plot reveals a massive spike in churn during months 0–12. If a customer survives the first year, they are likely to stay for 4+ years.

---

4. Model Performance

We trained two models. **Logistic Regression** slightly outperformed Random Forest in this specific case, likely because the relationships are linear (e.g., lower tenure = higher churn).

Metric	Logistic Regression	Random Forest
Accuracy	81%	80%
ROC-AUC	0.84	0.84
Recall (Churners)	56%	52%

- **Top Predictors (Feature Importance):**
    1. **Contract (Month-to-Month):** The #1 risk factor.
    2. **Tenure:** New customers are the most vulnerable.
    3. **TotalCharges:** High lifetime value correlates with retention.
    4. **InternetService (Fiber Optic):** Surprisingly high churn, indicating potential dissatisfaction with price or reliability.

---

5. Postdictive Analysis (The "Why")

The final_churn_prediction_analysis.csv file compares our predictions against reality.

- **Where we won:** The model perfectly flags "Price Sensitive Hoppers" - customers with high monthly bills, month-to-month contracts, and low tenure.
- **Where we lost:** The model missed "Long-Term Sleepers" - loyal customers of 5+ years who suddenly left. This suggests we need "Event Data" (e.g., a recent service outage) to catch them.

---

6. Business Recommendations

1. **The "Bundle & Save" Strategy:** Since churn drops with TotalServices, aggressive upsell campaigns should target 0-1 service customers. "Get Online Backup for $1" is a retention play, not just revenue.
2. **The "Year-One" Bridge:** Implement a "3-Month Check-in" for new customers. Churn is highest in year 1; surviving this period creates long-term value.
3. **Fiber Optic Audit:** Fiber users churn more than DSL users despite faster speeds. Investigate if this is due to technical outages or aggressive competitor pricing in fiber zones.